603 Group Project Intro

Data 603 Group [IDK what our group number is]

2024-11-30

Predicting Superconductor Critical Temperatures with Multivariate Regression

Alina Yildir UCID:

Archana Senthil UCID:

Gurmol Sohi UCID:

Trevor Yonkman UCID:

William Hovdestad UCID: 30248416

---

1. INTRODUCTION

1.1 MOTIVATION

1.1.1 Context

Certain materials exhibit the property of superconductivity, meaning that, under certain conditions, they can carry an electrical current with no losses. Since the discovery of the first superconductor in 1911 [1], which showed these properties in mercury below a Critical Temperature (Tc) of 4.2 Kelvin (-269 °C), researchers have looked for materials with superconducting properties at higher and higher temperatures. Applications of superconductors include medical imaging [2] and quantum computing [3], with each becoming significantly cheaper and more scalable as more favorable materials are found.

1.1.2 Problem

By modeling data from known materials, this report will attempt to identify parameters of significance leading to higher values of Tc. This may provide some clues for the properties that researchers should look for in newly proposed materials. There is some research precedent investigating new materials using statistical methods, with some success [4]. Methods carried out in prior research include linear regression, as well as more complex machine learning techniques like artificial neural networks and gradient boosting decision trees [5].

1.1.3 Challenges

One recent well-known example which speaks to the difficulty of discovering these materials was proposed by researchers [6] at Korea University in July 2023. This material, a grey-black, polycrystalline compound, identified as a copper-doped lead-oxyapatite and named LK-99 by the team which discovered it, was claimed to exhibit superconducting properties at room temperature and ambient pressure. While the results were never replicated, the months following LK-99's publicity saw a flurry of media reporting on the great significance of what this discovery would mean for new technologies [7].

Challenges specific to discovering new materials through linear regression include:

1. Non-linear relationships
   Superconductivity is governed by complex physical phenomena that often exhibit non-linear relationships. Linear regression assumes a linear relationship between variables, which may not capture these complexities accurately.
2. High dimensionality
   Superconducting materials are characterized by numerous properties, such as first ionization energy, atomic radius, electron affinity, and fusion heat. This high-dimensional data must be handled carefully to avoid overfitting.
3. Limited data availability.
   Experimental data on superconducting materials is scarce and often noisy. Linear regression models require a large amount of data to produce reliable predictions, and limited data can lead to inaccurate models.

1.2 OBJECTIVES

1.2.1 Overview

The overall intent of the project is to create a regression model to predict the Critical Temperature (Tc) of different materials. An accurate regression model to predict Tc can be used to identify which factors lead to a higher Tc. However, great care must be taken if looking for Tc values above the range of values used in creating our model.

1.2.2 Goals & Research Questions

What combination of factors are the most accurate for predicting the Critical Temperature (Tc) of a given material?

What factors should a material scientist focus on when looking for new superconductor materials?

Can we use the measured properties of existing materials to accurately predict a new (hypotheical) superconductor material with higher Tc then currently existing materials?

---

2. METHODOLOGY

2.1 DATA

2.1.1 Dataset Summary

The data-set we imported and analysed in R in a CSV file called "train.csv". It's 20,000 rows contains various different measurements of various different atomic properties of various materials, the number of elements this material contains, and finally the critical temperature (in Kelvin) that we are trying to create a regression model for.

2.1.2 Dataset Source

The dataset we obtained was sourced from the github repository of the "A data-driven statistical model for predicting the critical temperature of a superconductor" [8] paper. The github repository for the paper includes pre-processed data that has been cleaned and is ready for statistical analysis, as well as it's own analysis of said data.

We initially planned to start with the same raw data as the author of this paper; however, this 2018 paper cites a now-defunct website as the source for their raw-data: http://supercon.nims.go.jp/supercon/material_menu. Attempts were made to find an updated source of the raw data, but we were unable to locate an easily accessible and readable version of the raw data. Given our limited timeline, and the availability of data that had been pre-processed and was ready for statistical analysis, we chose to go-ahead with the pre-processed data.

2.1.2 Dataset Variables Examination

Each row of the dataset represents various different measurements and/or properties from a unique material; each distinct measurement and/or property is

The variables of our dataset are made up of 82 different measurements and/or properties from a unique material. While initially daunting, we can breakdown the variables into categories.

The first variable (column 1) is `number_of_elements` which represents the number of elements in the distinct material.

The last variable (column 82) is `critical_temp` which represents the critical temperature of the material, which we are trying to create a regression formula to predict.

The remaining 80 variables (column 2 through column 81, inclusive) represent different measurements of atomic properties. These 80 columns can be broken down into 8 ***atomic properties***, with 10 ***measurements*** per atomic property.

Atomic Properties (8 total)

1. Atomic Mass: the quantity of matter contained in an atom of an element. The observed atomic mass is slightly less than the sum of the mass of the protons, neutrons, and electrons that make up the atom.
2. First Ionization Energy: is the energy needed to remove the outermost, or highest energy, electron from a neutral atom in the gas phase.
3. Atomic Radius: the average distance from the center of an atom's nucleus to its outermost electron shell.
4. Density: a characteristic property of a substance that measures how much mass is in a given volume. It's calculated by dividing the mass of a substance by its volume ($D = m/v$)
5. Electron Affinity: a measure of the attraction between the incoming electron and the nucleus - the stronger the attraction, the more energy is released.

6. Fusion Heat: the amount of heat energy required to change a solid substance into a liquid at its melting point.
7. Thermal Conductivity: the ability of a given material to conduct/transfer heat. It is generally denoted by the symbol K but can also be denoted by λ
8. Valence: A whole number that represents the ability of an atom or a group of atoms to combine with other atoms or groups of atoms.

Measurements (10 total)

1. Mean: an average value in math.
2. Weighted Mean: also known as a weighted average, is a type of average where each item in the set has a weight or frequency assigned to it, rather than being treated equally.
3. Geometric Mean: a mean or average, defined as the nth root of the product of the n values for the set of numbers.
4. Weighted Geometric Mean: a generalization of the geometric mean using the weighted arithmetic mean, given a sample and weights.
5. Entropy: a measure of the amount of energy which is unavailable to do work.
6. Weighted Entropy: the measure of information supplied by a probabilistic experiment whose elementary events are characterized both by their objective probabilities and by some qualitative (objective or subjective) weights.
7. Range: the area of variation between upper and lower limits on a particular scale.
8. Weighted Range: used to define the quantity requirement of a pattern definition. It is used to express quantity related conditions like: exactly 5, or at least 5.
9. Standard Deviation: a measure of how dispersed the data is in relation to the mean.
10. Weighted Standard Deviation: allows you to apply a weight, or relative significance to each value in a set of values.

2.1.3 Full List of Variables

Explanatory variables:

1. `number_of_elements`: Number of distinct elements in the material.
2. `mean_atomic_mass`: Average atomic mass of elements.
3. `wtd_mean_atomic_mass`: Weighted average atomic mass, considering each element's proportion.
4. `gmean_atomic_mass`: Geometric mean of atomic mass across elements.
5. `wtd_gmean_atomic_mass`: Weighted geometric mean of atomic mass.
6. `entropy_atomic_mass`: Entropy (diversity) of atomic mass in the compound.
7. `wtd_entropy_atomic_mass`: Weighted entropy of atomic mass.
8. `range_atomic_mass`: Range of atomic masses in the compound.
9. `wtd_range_atomic_mass`: Weighted range of atomic mass.
10. `std_atomic_mass`: Standard deviation of atomic mass.
11. `wtd_std_atomic_mass`: Weighted standard deviation of atomic mass.
12. `mean_fie`: Mean first ionization energy of elements. Ionization energy affects electron mobility, which is essential for superconductivity.
13. `wtd_mean_fie`: Weighted mean first ionization energy.
14. `gmean_fie`: Geometric mean of first ionization energy. Indicates central ionization tendency, relevant for electron stability.

15. wtd_gmean_fie: Weighted geometric mean of ionization energy. Reflects effective ionization tendency based on element proportion, impacting superconductivity.
16. entropy_fie: Entropy of ionization energy, showing diversity in electron binding energy.
17. wtd_entropy_fie: Weighted entropy of ionization energy. Gives a more precise diversity measure, impacting electron mobility for superconductivity.
18. range_fie: Range of ionization energies. Greater range implies variation in electron release energy, influencing Tc.
19. wtd_range_fie: Weighted range of ionization energy. Effective diversity measure, impacting electron stability and superconducting behavior.
20. std_fie: Standard deviation of first ionization energy.
21. wtd_std_fie: Weighted standard deviation of first ionization energy.
22. mean_atomic_radius: Average atomic radius of elements.
23. wtd_mean_atomic_radius: Weighted mean atomic radius.
24. gmean_atomic_radius: Geometric mean of atomic radius. Indicates central size tendency, relevant to lattice dynamics.
25. wtd_gmean_atomic_radius: Weighted geometric mean of atomic radius.
26. entropy_atomic_radius: Entropy of atomic radius, showing size diversity.
27. wtd_entropy_atomic_radius: Weighted entropy of atomic radius.
28. range_atomic_radius: Range of atomic radii in the compound.
29. wtd_range_atomic_radius: Weighted range of atomic radius.
30. std_atomic_radius: Standard deviation of atomic radius. High variation impacts stability and electron dynamics relevant for superconductivity.
31. wtd_std_atomic_radius: Weighted standard deviation of atomic radius.
32. mean_Density: Average density of the material.
33. wtd_mean_Density: Weighted mean density, adjusted for element proportions.
34. gmean_Density: Geometric mean of density.
35. wtd_gmean_Density: Weighted geometric mean density.
36. entropy_Density: Entropy of density, showing diversity in packing.
37. wtd_entropy_Density: Weighted entropy of density.
38. range_Density: Range of density values.
39. wtd_range_Density: Weighted range of density.
40. std_Density: Standard deviation of density.
41. ‘wtd_std_Density: Weighted standard deviation of density.
42. mean_ElectronAffinity: Average electron affinity of elements.
43. wtd_mean_ElectronAffinity: Weighted electron affinity, adjusted for composition.
44. gmean_ElectronAffinity: Geometric mean of electron affinity.
45. wtd_gmean_ElectronAffinity: Weighted geometric mean electron affinity.
46. entropy_ElectronAffinity: Entropy of electron affinity, showing diversity in attraction potential.
47. wtd_entropy_ElectronAffinity: Weighted entropy of electron affinity.
48. range_ElectronAffinity: Range of electron affinity values.
49. wtd_range_ElectronAffinity: Weighted range of electron affinity.
50. std_ElectronAffinity: Standard deviation of electron affinity.
51. wtd_std_ElectronAffinity: Weighted standard deviation of electron affinity.
52. mean_FusionHeat: Average fusion heat of elements.
53. ‘wtd_mean_FusionHeat: Weighted fusion heat, adjusted for element composition.
54. gmean_FusionHeat: Geometric mean of fusion heat.
55. wtd_gmean_FusionHeat: Weighted geometric mean fusion heat.
56. entropy_FusionHeat: Entropy of fusion heat, showing melting energy diversity.
57. wtd_entropy_FusionHeat: Weighted entropy of fusion heat.
58. range_FusionHeat: Range of fusion heat.
59. wtd_range_FusionHeat: Weighted range of fusion heat.
60. std_FusionHeat: Standard deviation of fusion heat.
61. wtd_std_FusionHeat: Weighted standard deviation of fusion heat.
62. mean_ThermalConductivity: Average thermal conductivity of the material.

63. `wtd_mean_ThermalConductivity`: Weighted mean thermal conductivity.
64. `gmean_ThermalConductivity`: Geometric mean of thermal conductivity.
65. `wtd_gmean_ThermalConductivity`: Weighted geometric mean of thermal conductivity.
66. `entropy_ThermalConductivity`: Entropy of thermal conductivity, showing diversity in heat transfer.
67. `wtd_entropy_ThermalConductivity`: Weighted entropy of thermal conductivity.
68. `range_ThermalConductivity`: Range of thermal conductivity values.
69. `wtd_range_ThermalConductivity`: Weighted range of thermal conductivity.
70. `std_ThermalConductivity`: Standard deviation of thermal conductivity.
71. `wtd_std_ThermalConductivity`: Weighted standard deviation of thermal conductivity.
72. `mean_Valence`: Average valence electron count.
73. `wtd_mean_Valence`: Weighted average valence electrons.
74. `gmean_Valence`: Geometric mean of valence electrons.
75. `wtd_gmean_Valence`: Weighted geometric mean of valence electrons.
76. `entropy_Valence`: Entropy of valence electron distribution, showing bonding diversity.
77. `wtd_entropy_Valence`: Weighted entropy of valence electrons.
78. `range_Valence`: Range of valence electrons.
79. `wtd_range_Valence`: Weighted range of valence electrons.
80. `std_Valence`: Standard deviation of valence electrons.
81. `wtd_std_Valence`: Weighted standard deviation of valence electrons.

Derived variable:

82. `critical_temp`: The Critical Temperature (Tc), the critical temperature below which a given material acts as a superconductor.

2.2 APPROACH

What approach are you using for data/visual analytics solution? Why do you think it will work well?

2.3 WORKFLOW

What steps (workflow task list) are required? Which of these steps is particularly hard? What to do if the hard steps don't work out

---

2.4 CONTRIBUTIONS

Briefly describe the group members' workload distribution and responsibilities

---

3 MAIN RESULTS OF THE ANALYSIS

3.1 Results

What do my results indicate? Do you have any unexpected results? Please elaborate.

---

4 CONCLUSION AND DISCUSSION

4.1 Approach

Overall, is the approach we took promising? Please elaborate. What different approach or variant of this approach is better?

4.2 Future Work

What should follow-up work be done next?

---

5 REFERENCES

References

1. CERN. (n.d.) Superconductivity. Retrieved from https://home.cern/science/engineering/superconductivity [Accessed: Nov 8, 2024]
2. Questions and Answers in MRI. (n.d.) Superconducting Magnets. Retrieved from https://mriquestions.com/superconductive-design.html. [Accessed: Nov 8, 2024]
3. Pennylane. (n.d.) Quantum Computing with Superconducting Qubits. Retrieved from https://pennylane.ai/qml/demos/tutorial_sc_qubits/. [Accessed: Nov 8, 2024]
4. SN Applied Sciences (n.d.) Prediction of critical temperature and new superconducting materials | Discover Applied Sciences. Retrieved from https://link.springer.com/article/10.1007/s42452-020-03266-0?form=MG0AV3 [Accessed: Nov 29, 2024]
5. MDPI Open Access Journals (n.d.) Predictive Modeling of Critical Temperatures in Superconducting Materials. Retrieved from https://www.mdpi.com/1420-3049/26/1/8?form=MG0AV3 [Accessed: Nov 29, 2024]
6. Cornell University (n.d.) The First Room-Temperature Ambient-Pressure Superconductor. Retrieved from https://arxiv.org/abs/2307.12008 [Accessed: Nov 29, 2024]
7. CBC (n.d.) Claims of revolutionary superconductor LK-99 are meeting resistance. Here's what you need to know. Retrieved from https://www.cbc.ca/news/science/lk99-superconductor-resistance-explainer-1.6929966 [Accessed: Nov 29, 2024]
8. K. Hamidieh, "A data-driven statistical model for predicting the critical temperature of a superconductor," Computational Materials Science, vol. 154, pp. 346–354, Nov. 2018. DOI: https://doi.org/10.1016/j.commatsci.2018.07.052.

Appendix

Optional part of the proposal. Include any support material you may have, e.g., figures, tables, samples of your data set(s). Make sure to cite any of the Appendix material in the main proposal text.